IN THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF NEW YORK

HACHETTE BOOK GROUP, INC., HARPERCOLLINS PUBLISHERS LLC, JOHN WILEY & SONS, INC., and PENGUIN RANDOM HOUSE LLC Plaintiffs, v. INTERNET ARCHIVE and DOES 1 through 5, inclusive Defendants.	Case No. 1:20-CV-04160-JGK

EXPERT REPORT OF IMKE REIMERS, PH.D.

ATTORNEYS' EYES ONLY

TABLE OF CONTENTS

Page

I. QUALIFICATIONS AND STATEMENT OF ASSIGNMENT

1. I am an associate professor of economics at Northeastern University in Boston, Massachusetts ("Northeastern"). I began working at Northeastern in 2014 as an assistant professor, and I was promoted to an associate professor (with tenure) in 2020. At Northeastern University, I have been teaching undergraduate economics classes in microeconomic theory and applied econometrics as well as a Ph.D.-level course in which I introduce students to the frameworks of industrial organization. In addition, I have supervised several Ph.D. students in writing their dissertations.

2. Prior to joining Northeastern, I received my Bachelor of Science degree in mathematics and economics in 2008 from the University of Nebraska, my Master of Arts degree in economics from the University of Minnesota in 2012, and my Ph.D. in economics, also at the University of Minnesota, in 2013. After finishing my Ph.D. and before joining Northeastern, I spent one year as the Economics of Digitization post-doctoral fellow at the National Bureau of Economic Research ("NBER"), where I participated in the planning of activities for the NBER Digitization and Copyright Initiative and helped curate a website aimed at improving access to datasets related to the economics of digitization.

3. My main area of research is industrial economics with a focus on digital markets and intellectual property. More specifically, I have written and published six papers on various topics around book publishing in peer-reviewed economics and management science journals, as well as four other papers on digital markets and enforcement of intellectual property rights—especially copyrights. Much of my research involves studying how digitization has changed access to information about products, firms, and consumers, and how these changes affect the profitability of firms as well as the well-being of consumers—the ability of readers to purchase and read books they enjoy. For example, in my dissertation, I studied the impact of the 1998

Copyright Term Extension Act on the availability of books that were affected by the extension.[1] I found that books that continued to be protected by copyright were significantly less likely to be available than comparable books that had moved into the public domain, and that potential gains to copyright holders were not large enough to offset losses to consumers. In other work on the book publishing industry, I found that piracy takedown notices that were directed at individual or small-scale infringers were effective in raising ebook sales.[2] In more recently published works, I focused on the role that information plays for both publishers and consumers. In one paper, I found that improvements in the information environment for potential readers—via Amazon star ratings–help them make better purchasing decisions.[3] In another paper, I found that publishers' advances to authors have begun to better reflect an author's ex-post success in terms of book sales for genres that are most affected by self-publishing, suggesting that publishers have also been able to benefit from the information made available via digital markets.[4] I continue to work on topics around book publishing, including studies of the role impacts of Google Books and the changing role of libraries. My curriculum vitae, which includes a list of publications I authored, is attached as **Exhibit A**.

---

[1] See Reimers, Imke. "Copyright and generic entry in book publishing." *American Economic Journal: Microeconomics* 11.3 (2019): 257-84.

[2] See Reimers, Imke. "Can private copyright protection be effective? Evidence from book publishing." *The journal of law and economics* 59.2 (2016): 411-440.

[3] See Reimers, Imke, & Joel Waldfogel. "Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings." *American Economic Review* 111.6 (2021): 1944-71.

[4] See Peukert, Christian & Imke Reimers. "Digitization, Prediction, and Market Efficiency: Evidence from Book Publishing Deals." *Management Science* (2022).

4. I have not testified under oath on any matter in the past—either at trial or at deposition. I am being paid $300 per hour for my work on this case. Payment is not contingent on my opinion expressed.

5. Apart from the drafting of this expert report, I do not have a connection or affiliation with Plaintiffs or Defendant, the Internet Archive.

6. I have been asked by counsel for the Internet Archive to evaluate and describe the following:

a. The economic performance and growth of the publishing industry while the Internet Archive's digital lending program was in operation—especially (i) from 2017 to June 2020, the month in which this lawsuit was initiated, and (ii) when the National Emergency Library ("NEL") was in effect (March 24, 2020 through June 16, 2020);

b. The commercial life cycle of a book relative to the date of publication; and

c. The relationship between a title's availability through the Internet Archive's digital lending program and that title's commercial performance, with a particular focus on older books—those on the backlist.

7. This report is based on my analysis of the materials I have reviewed to date. Specifically, I have collected granular sales ranking data from Amazon's website for the set of titles listed in the Complaint, I have reviewed academic and trade publications, and I am drawing on research and data that I have worked on as an independent researcher prior to this case. My work in this matter is ongoing, and I reserve the right to supplement my opinions and analysis should any additional information become available to me. I also intend to review any additional information or reports that may be submitted by Plaintiffs and their experts, and I reserve the right to prepare a reply expert report. I have also reviewed documents provided to me by counsel

for the Internet Archive. A list of the materials I have considered in reaching my opinions is attached to this expert report as **Exhibit B**.

8. I understand that this case concerns Plaintiffs' allegations of copyright infringement against the Internet Archive, including Plaintiffs' contention that the Internet Archive's digital lending library depresses their revenues from the 127 books listed in the complaint (the "Works in Suit"). I also understand that two systems of digitized lending are being challenged in this lawsuit. One is Controlled Digital Lending as practiced by the Internet Archive's digital lending library. And the other is the NEL. Controlled Digital Lending as I understand it is a system of digitizing physical books and loaning those digitized versions out to patrons on an owned-to-loaned ratio. In other words, if more patrons want at a given time to borrow a digitization than the number of physical copies an institution owns and has included in the digital lending program, the excess number of patrons must wait until a digitized copy is checked in before accessing the book. I understand that the NEL, which was operational for approximately three months in 2020 as a response to the COVID-19 pandemic, did not enforce the owned-to-loaned ratio and that while the NEL was in effect, patrons did not have to wait to borrow a book.

II. SUMMARY OF CONCLUSIONS

9. I have reached the following conclusions:

a. Ebook sales have decreased steadily over the last five to ten years, with no evidence that the decrease in sales and revenues was spurred by the operation of the Internet Archive's digital lending program. Though there has been considerable growth of ebook sales starting in 2008, this growth has halted around 2013, and ebook sales declined from just under 473 million units in 2014 to just under 336 million units by 2019, according to the Association of

American Publishers ("AAP").[5] Print book sales, on the other hand, have remained relatively stable since 2014. Consequently, the share of ebook sales among all book sales has decreased steadily since 2014, from about 17.5% in 2014 to about 12% in 2019.

b. A large proportion of book sales—for both print and ebooks—occurs in the first few years after a title's publication. I provide information on the sales decay for relatively successful titles—those that reach the weekly top 150 USA Today bestseller lists.[6] Between 1994 and 2021, of all book appearances on the weekly bestseller lists, only 21% are of books that first appeared at least 52 weeks earlier, only 8% include books that debuted on the lists more than two years earlier, and only 2% of the listings are for books that are more than five years old. While this is indirect evidence, direct evidence from the publishers suggests similar patterns. HarperCollins, Penguin Random House, and Hachette report that between 33% and 60% of their unit sales during the period for which annual data was provided.[7] Because only a small fraction of each publisher's catalog was published in the past year, the share of each individual book's revenues that occurs in the first year is likely much larger than the share of the sales across the catalog that is attributed to books that are less than one year old. In addition, annual sales of the Works in Suit suggest similar decreases in a book's lifetime sales, where sales in the first five years after an edition's publication account for up to 90% of lifetime sales.

[5] Annual AAP StatShot Reports (AAP001108; AAP001131).

[6] See, for example, "Best-Selling Books," *USA Today*, available at https://www.usatoday.com/entertainment/books/best-selling/week/2021/1/page/1/.

[7] Data are from HC0030132, PRH0072194, and HACHETTE0012377. The reasoning for using data from Hachette, HarperCollins, and Penguin Random House for this observation—but not Wiley—is explained more fully below.

10. Based on my analysis of the Works in Suit, I have reached the following conclusions:

a. I find no evidence that availability of these titles for borrowing from the Internet Archive's digital lending program depressed book sales of print books through other channels.[8] In particular, I find no evidence that sales rankings for print books of the Works in Suit at Amazon worsened (as in, moved away from a ranking of 1) when the works were made available through the Internet Archive. I also find no evidence that the launch of the NEL depressed sales.

b. By contrast, I find some (weak) evidence that taking the Works in Suit off the Internet Archive hurt the Amazon sales rankings of their print editions slightly. This finding suggests that physical embodiments of the Works in Suit had lower sales after they were removed from the Internet Archive, relative to other books on Amazon and compared to the weeks before the books were taken down.

III. DATA

11. To conduct my analyses, I draw on data from several sources. In addition to industry sales figures obtained through trade publications, deposition transcripts and exhibits to those depositions, and documents produced by the AAP, I collected (i) weekly top 150 book bestseller lists from 1994 to 2021 through USA Today, and (ii) daily Amazon sales rankings and

[8] As I explain below, I focus my analysis on print books because title-specific data for ebook sales were not available at a granular enough level to enable me to draw reliable conclusions. In addition, and equally important, granular ebook sales data are not available for other titles, which were not available at the Internet Archive, so there is no proper group of titles to compare changes in sales. Because book readership may have changed for other reasons over time, this makes it difficult to establish a causal relationship between availability at the Internet Archive and sales through other channels. Still, an analysis that focuses on physical book sales is relevant because print book sales continue to make up a large majority of unit sales and revenue for most books, as I summarize in Paragraph 10a and lay out in more detail in Paragraphs 34 – 54.

other product characteristics for 1,260 print editions of the Works in Suit through the website Keepa (www.keepa.com). This website (Keepa) provides application programming interface ("API") access to histories of daily sales rankings, prices, star ratings, and marketplace availability for specific books. In order to collect this information from Keepa, one has to search for specific book identifiers (either the ISBN-13 or the ISBN-10).[9] I provide information on all additional sources and details on data collection below.

IV. ANALYSIS

A. The Book Publishing Industry in Recent Years

12. The arrival of ebooks substantially changed the market for books. Because of ebooks, people interested in reading a book no longer had to physically visit a bookstore or wait for a book to be delivered from an online retailer. Ebooks enabled instant gratification: readers could choose a book online and immediately download it to their ereading device. This had the potential to change how books were primarily purchased.

13. In addition, the arrival and rise of self-publishing platforms such as Kindle Direct Publishing (2007, concurrent with the introduction of the Kindle ereader) and Smashwords (2008) posed a threat to publishers by providing authors a direct channel to reach potential readers. However, annual sales data from the AAP suggest that traditional publishers were able to keep their sales high throughout this period of change.[10]

[9] This website has been used in a series of academic works, including two papers that I have written. For instance, I have used this website in a paper on the effects of Amazon star ratings and professional book reviews on book sales, which was published in 2021 in the American Economic Review.

[10] Annual AAP StatShot Reports (AAP001108; AAP001131; AAP000804).

14. The 2007 introduction of Amazon's Kindle ereader made ebooks a viable alternative to their physical counterparts. Although other ereading devices had been available to consumers earlier, the Kindle ereader presented a much-improved reading experience due to its use of E Ink technology, which produces a paper-like display that requires very little battery power because it does not require the screen to be backlit. Moreover, Amazon's Kindle Store made it easy to download ebooks straight to the Kindle after purchase.

15. From 2008 to 2013, ebook sales reached about 242 million units.[11] According to the annual AAP StatShot reports, total ebook unit sales decreased steeply from 2014 to 2016, from 472.7 million units (and $3.2 billion in revenue) to 366.1 million units ($2.2 billion in revenue).[12] After 2016, the rate at which ebook sales were declining slowed but the decline did not stop until 2020: ebook unit sales continued to decrease over the next three years, landing at 335.7 million units ($1.94 billion in revenue) in 2019. In 2020, ebook revenue increased to $2.12 billion.[13] From 2016 to 2020, revenue from ebooks decreased by 8.5%.

16. Sales of physical books have changed coinciding with the arrival of ebooks. Between 2008 (the arrival of the Kindle ereader) and 2012, annual sales of print books fell from

[11] I am not aware of any resources that track ebook sales prior to 2008. Ebook unit sales for the years 2008 to 2013 are obtained from *Statista*, available at https://www.statista.com/statistics/426799/ebook-unit-sales-usa/, which reports data from Nielsen, the NPD Group, and Publishers Weekly. Ebook unit sales and revenues for the years 2014 to 2019 are obtained from the AAP Annual StatShot Reports for 2018 and 2019, as produced in this lawsuit by the AAP (AAP001108; AAP001131). The AAP reports significantly higher levels of ebook sales than Statista, likely because the reporting methodology differs from secondary sources like Nielsen, the NPD Group, and Publishers Weekly. For example, the number/identity of the reporting publishers or the included book categories (e.g., trade books, higher education) may vary across these sources.

[12] AAP001108; AAP001131.

[13] AAP000804 at 5. The 2020 AAP Annual StatShot report lists revenues but not industry unit sales by format.

778 million to 591 million units.[14] This suggests that some consumers may have begun to buy ebooks instead of physical books, and the two formats may therefore be substitutes for one another.

17. According to the Annual AAP StatShot reports from 2019 and 2020,[15] which contain information for print format sales across all book categories from 2014 to 2019, whereas ebook sales decreased over this time period, print sales remained either flat or increased slightly. Annual hardcover book sales, for instance, increased from 548.7 million units ($5.12 billion in revenue) in 2014 to 614 million units ($5.94 billion in revenue) in 2019, and revenue from hardcover books increased sharply in 2020, to $6.31 billion. Considering hardcover, paperback, and mass market paperback formats of physical books together, physical unit sales decreased slightly over the same period, from 2.30 billion units in 2014 to 2.17 billion units in 2019, although revenue *increased* slightly over this period, from $11.2 billion in 2014 to $12.0 billion in 2019.

18. Two conclusions emerge from the above analysis. First, any potential expectations of ebook sales surpassing print sales did not come to fruition by 2020. Ebook sales made up less than 20% of the total unit sales between 2014 and 2019: the share of unit sales that was attributed to ebooks was 17.5% in 2014 and decreased to 12.1% in 2019. Because ebooks tend to be less expensive than print books, the share of *revenue* from ebooks is even smaller than

[14] Print book unit sales for the years 2008 to 2013 are obtained from *Statista*, available at https://www.statista.com/statistics/422595/print-book-sales-usa/, which reports data from Nielsen and NPD Bookscan as well as Publishers Weekly and purports to include data for about 85% of physical book sales. From the source, it is unclear whether these data include all print sales or only certain formats and categories. For example, it is unclear whether these data are limited to trade books or also include higher education books.

[15] AAP001108; AAP001131; AAP000804.

ebooks' share of unit sales, decreasing from 11.6% in 2014 to 7.5% in 2019, and rising to 8.2% in 2020. Over the same period, the share of hardcover, paperback, and mass market unit sales among all publishing sales fluctuated between 62% and 68%, and their share of revenues increased from 40% in 2014 to 46% in 2019 and 48% in 2020.[16]

19. Second, whereas sales of ebooks and sales of print books have fluctuated relative to each other between 2008 and 2019, total sales across physical and ebook formats remained roughly the same over this period, with a slight downward trend. The AAP StatShot reports show that total unit sales across hardcover, paperback, mass market, and ebook formats decreased by 7.7% from 2014 (2.23 billion units) to 2019 (2.06 billion units).

B. The Life Cycle of a Book's Sales

20. Analysis of any possible effect the Internet Archive might have on print or ebook sales needs to incorporate the timing of the digitization of a title relative to its original publication. The vast majority of titles are digitized and made available for borrowing from the Internet Archive's lending library at least five years after their original publication date.[17] As analysis in the balance of this section shows, peak sales for most titles occur within one to two years after the date of first publication. Accordingly, when most works become available for borrowing from the Internet Archive's lending library, they are well past the peak of their sales.

21. I examined the life cycle of a book's sales in three ways. First, I examined the weekly Top 150 best-selling books for all weeks from 1994 to 2021, as reported by *USA*

[16] According to the annual AAP StatShot reports, instructional material accounts for the vast majority of the remaining format sales.

[17] See Internet Archive notes about book collections and availability, INTARC00151319.

Today.[18] These lists are helpful because they are based on sales of both print and electronic editions of a title and because they allow for a detailed (weekly) breakdown of title ages. Second, I supplement this analysis with information provided by three of the publishers that breaks down unit sales and revenues for the catalog of front list titles (as defined by each publisher) and backlist titles (as defined by each publisher).[19] Third, I examine the annual change in unit sales for the Works in Suit to document the distribution of specific titles' sales patterns across multiple years.

22. Because I do not have access to granular sales data for a large set of individual titles over a long period of time, I collected a long series of the weekly Top 150 best-selling titles, from 1994 to 2021, from USA Today's bestseller lists. These lists allow me to determine the approximate ages of each title on the lists.[20] In turn, this allows me to examine how long

[18] See, for example, "Best-Selling Books," *USA Today*, available at https://www.usatoday.com/entertainment/books/best-selling/week/2021/1/page/1/.

[19] I did not see catalog-wide revenue data separately for front list and backlist catalogs for Wiley. There is some variation in the way Plaintiff publishers define backlist. For instance, Hachette defines backlist as books published two or more years ago, front list as books published within one year, and prior year published as books published between the front list and backlist terms. Sevier Dep. Tr. 63:4–67:3. Penguin Random House, Wiley, and HarperCollins define backlist as books over a year old and front list as books younger than one year. Weber Dep. Tr. 101:21–104:9; Pavese Dep. Tr. 215:21–217:9; Restivo-Alessi Dep. Tr. 56:10–13. To avoid confusion, I refer not to front list or backlist but to the time relative to publication. Revenue information broken down by front list and backlist titles was available for HarperCollins and Hachette.

[20] I approximate each title's age by the time (in weeks) since its first appearance on the USA Today bestseller list. I do this because the title's original publication date is not listed in the USA Today bestseller lists and would therefore have to be searched for each individual title. A spot check of the true publication dates of all 150 bestsellers in one week (March 20, 2020) suggests that this is a reasonable approximation: the correlation between ranking debut dates and original publication dates is 0.83, which means that the ranking debut dates and original publication dates are closely related and move similarly and in the same direction (for instance, if one increases, then the other likely also increases). If the correlation were 1, then I know exactly by how much the other variable will move and that the variables would move in the same way. A correlation of 0.83 suggests a relatively close but not perfect relationship.

books tend to keep their sales high enough to remain on the list. For illustration, in each year, I observe 7,800 bestseller entries: 52 weeks, times 150 entries per week. In 2004, 5,413 of those entries were originally published in 2012. This could be either because books are most popular right after their release, or because books published in 2012 were of a higher quality (and therefore more popular) than books from previous years. To determine the role of sales decay (rather than quality differences across publication years), I then follow how many entries in the following years are made up by books that were originally published in 2012. I find that between 12 and 24 months after their debut, only 1,049 of the entries are from the 2012 cohort—a decrease of over 80%. Five to six years after their debut, only 97 entries remain on the bestseller list, 2.2% of the original 5,413.

23. There is a similar decline in the number of bestseller entries over time across all cohorts (debut years). Focusing on debut years before 2016 to ensure that I observe at least five years of sales history for all books, I find that the number of entries for books that are 1–2 years old is only 23% as large as the number of entries for books that are less than one year old; and that ratio decreases rapidly as books continue to age: books between two and three years account for only 8.7% as many entries as those less than one year old. After five years, that ratio decreased to 2.7% of those under one year old. Figure 1 illustrates this average decrease in the number of bestseller appearances in the years after a book's first appearance. Unless sales for bestselling titles decline at a faster rate than those for non-bestsellers, and there is no evidence to suggest this is so, this large decrease is suggestive of rather rapid declines in sales within less

than five years after initial publication—which suggests that this pattern holds true for the Works in Suit.[21]

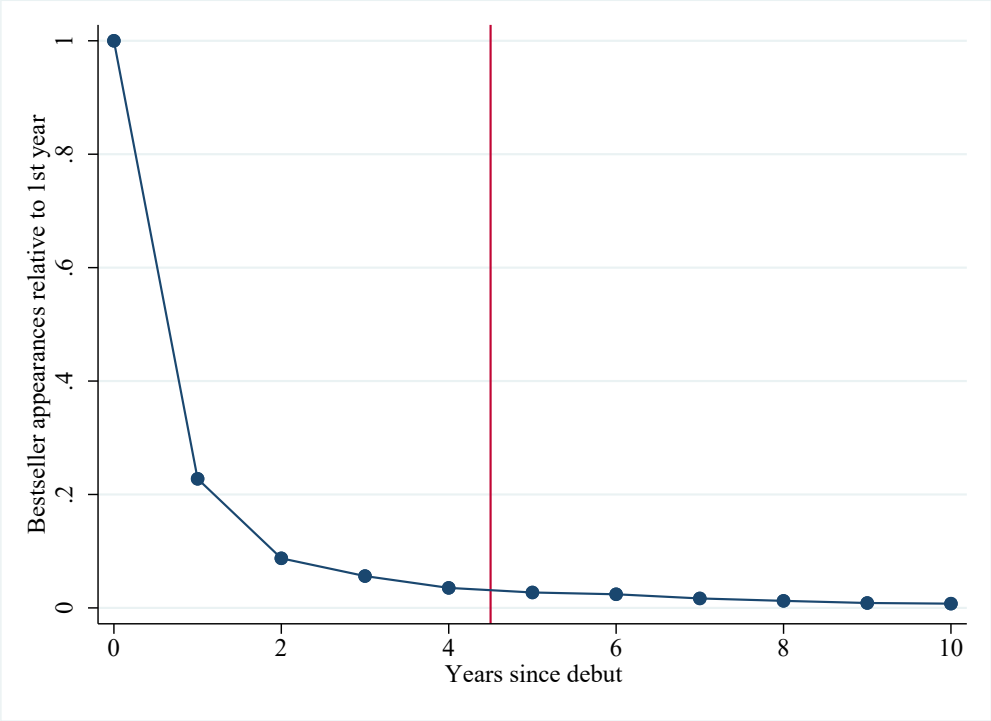

Figure 1: Average number of USA Today Top 150 Bestseller appearances per year since first appearance, relative to the first year

24. Paragraph 23 above shows averages across bestseller cohorts but does not directly account for the possibility that some cohorts are inherently more successful than others. In a more formal approach, I estimate a regression which controls for factors that make some cohorts more successful than others. This allows me to focus on the decay of bestseller entries for each

[21] Although I am not aware of evidence that the sales trajectories of bestsellers behave differently from those of non-bestsellers, I cannot rule this out. I provide evidence of similar sales decreases at the title level for the Works in Suit in Paragraph 27 below.

individual cohort.[22] The regression provides estimates of the number of bestseller entries in each year after a book's debut on the bestseller list. I illustrate the results in Figure 2, in which I plot the share of bestseller entries that have already happened (y axis) by the book's age (x axis). For example, the value of 0.62 in year 0 implies that—on average, and after accounting for differences in popularities—62% of a book's bestseller appearances happen within the first year. After four years, I find that 88% of all bestseller entries have already occurred, indicating that only 12% of the times that books appear on the bestseller lists occur more than five years after their debut.

[22] Formally, I estimate a fixed effects regression model. To do this, I create a dataset that summarizes the number of entries for each bestseller debut cohort in each year after their debut (in other words, the number of times a title appears on the bestseller list in the year following its publication). I then estimate, for each title the number of bestseller entries per year since the title's debut, with a set of indicator variables (fixed effects). The purpose of this analysis is to control for circumstances where certain titles (cohorts) have longer than normal popularity. For example, an indicator variable for debut year 2004 will take on the value 1 for observations that describe the debut year 2004 and a value of 0 for all other debut years. One can think of these fixed effects as separate y-intercepts for each debut year. This is what allows me to identify changes in the number of entries as books age.

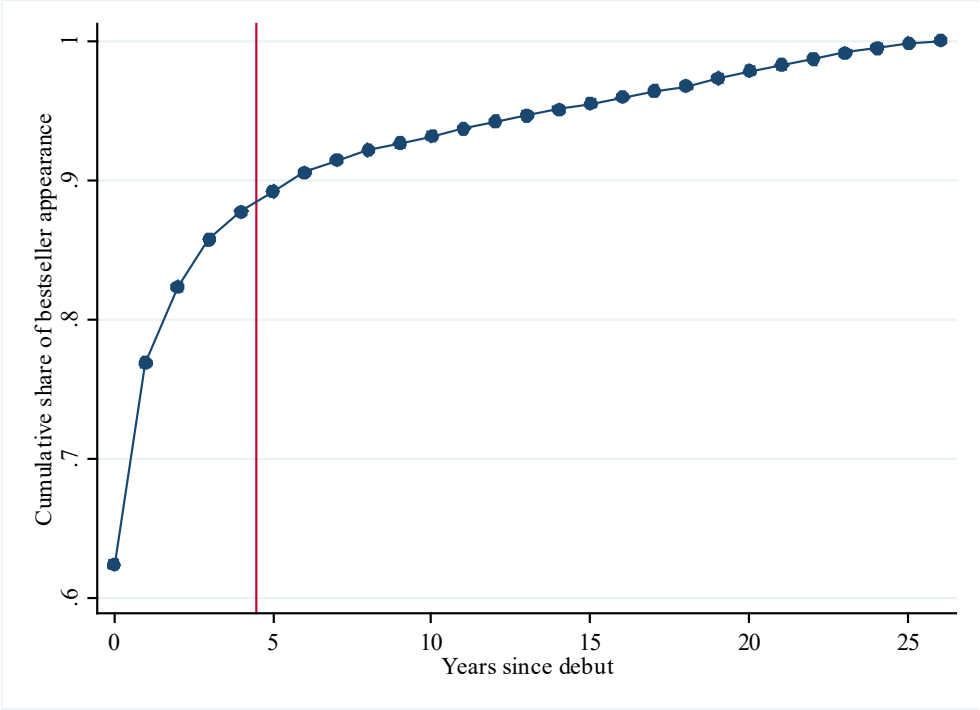

Figure 2: Cumulative share of USA Today bestseller list appearances by year since first appearance (for all bestselling books originally appearing before 2013).

25. Revenue reports made available by Plaintiffs are largely in line with the above analysis. For example, HarperCollins reports that sales of books that were published less than a year prior among print books accounted for between 37% and 40% of all units sold between 2017 and 2019, and this fraction dropped to 34% in 2020. [23] Penguin Random House reports that between 2017 and 2019 books published less than a year prior accounted for 52% to 47%, and 43% in 2020.[24] Hachette reports that between 2017 and 2019, the total share of revenues accounted for by titles published at most two years prior was between 70% and 73%, and the share fell to 65% in 2020.[25] At Hachette, the share of revenue relative to the life of a title that is

[23] See HC0030132.

[24] See PRH0072194.

[25] See HACHETTE0012377.

due to titles less than two years old is smaller for ebooks (around 60%) and larger for physical formats (74% to 77% in 2017 to 2019, and 69% in 2020). Note, though, that only a small fraction of each publisher's catalog was published in the past year—indeed a small fraction is ever published in a given year. Therefore, the share of each individual book's revenues that occurs in the first year is likely much larger than the share of a publisher's sales across its catalog that is attributed to books that are less than one year old.

26. I also observe annual unit sales and revenue for the Works in Suit. For the Works in Suit that are published by Penguin Random House, I observe sales for individual book editions from 2010 to 2020; for Hachette, I observe similar information for the years 2015 to 2019; for HarperCollins, I observe title and edition performance from 2017 to 2020; and for Wiley, I observe information for a wider range of years (from 1996 to 2021), but the data cover very few (less than ten) ISBNs before 2007.[26] For each of the editions for which I am able to observe all sales since their publication, I calculate the share of total sales (up to the year 2020) that occurred in each year since publication. For example, for each Penguin Random House book edition that is published in 2010, I calculate the total unit sales from 2010 to 2020, and I then calculate the share of these sales that occurred in 2010, 2011, and so on. [redacted]

27. [redacted]

[26] See PRH0025907; HACHETTE0002474; HC0010272; and WILEY0005650.

[redacted]

[redacted]

[redacted]

28. To summarize, my analyses suggest that—at least for titles that enter the Top 150 bestseller lists and for most of the Works in Suit—around 90% of their sales occur in the first five years after their publication.

C. The Role of Libraries and Their Interaction with Book Publishing

29. Libraries have traditionally played a large role in making print copies available to the public, and most public libraries have embraced ebook holdings and circulation as well. The Institute of Museum and Library Services ("IMLS") performs an annual survey of public libraries, in which it asks, among others, about the library's number of physical book and ebook holdings, as well as the number of instances of circulation of each format. Data from these annual Public Library Surveys ("PLS") are available at the IMLS website.[30] According to the 2019 PLS—the most recent available survey—library expenditures on print formats amounted to $497 million in 2019, and library expenditures on electronic content was reported at $257.1 million. Library expenditures on print formats remained quite stable between 2014 and 2019, ranging from $497 million to $507.5 million. Library expenditures on electronic formats, on the other hand, increased by over 50% over the same period, from $166.7 million in 2014 to $257.1 million in 2019.

30. Library expenditures increased in recent years, but library expenditures as a share of the total publishing market remain low. Combining data from the PLS with the annual AAP

[30] See "Public Libraries Survey," *IMLS*, available at https://www.imls.gov/research-evaluation/data-collection/public-libraries-survey. The survey for 2020 is still underway as of January 26, 2022.

StatShot reports, I find that library expenditures on physical formats accounted for only 4.1% to 4.5% of total publisher revenue from hardcover, paperback, and mass market formats between 2014 and 2019. The share of library expenditures among all ebook revenue is higher, ranging from 5.1% in 2014 to 13.3% in 2019.

31. There is little hard evidence on the effect of libraries on book sales. This is because a credible analysis of such a relationship needs data on both detailed library holdings and circulation as well as local sales data of the same book.[31] In addition, such an analysis requires that the holdings at the library change for reasons other than an (anticipated) increase in demand. For example, a library might decide to purchase copies of a book on the history of the Olympics shortly before the start of the Olympics because it expects interest for that topic and title to increase. Then, we might see a rise in sales for the book concurrent with an increase in its library holdings. However, this positive correlation would describe a trend in interest, rather than a causal effect of library holdings on book sales.

32. Finding a setting that avoids these title-level common trends is difficult, but changes in the availability of a title at the Internet Archive can provide such a setting. Titles may have been made available at the Internet Archive following the five-year-post-publication rule imposed by the Internet Archive rather than due to any expectations in demand. In addition, the Works in Suit were taken off the Internet Archive in response to takedown notices sent by or on behalf of Plaintiffs, and there is no evidence that readership for these works changed for other reasons at this time. I take advantage of this setting in Section D below.

[31] This is because patrons of a particular library are more likely to purchase books from retailers convenient to (which sometimes means physically close to) them.

33. To summarize, I find in this section that libraries make up a small fraction of the overall publishing market, but that the share of library expenditures for ebooks has increased in recent years.

D. The Effect of the Internet Archive on Print Book Sales

34. It is possible that changes in the availability of a title at the Internet Archive affected sales through other channels in a meaningful way. Any analysis that tries to quantify the effects of availability at the Internet Archive on sales of a title is complicated by the possibility of concomitant changes in demand for the particular book and for reading in general that are unrelated to the book's availability at the Internet Archive. This would be particularly evident in any analysis of the effect of the Internet Archive's NEL on book sales.

35. The NEL coincided with an unprecedented shock to the market for books. The COVID-19 pandemic may have increased the demand for books because people spent more time at home.[32] Therefore, one cannot simply look at absolute changes in sales. One way to circumvent this issue is to compare changes in sales for the Works in Suit to changes in sales of similar books. However, identification of such "similar" books is difficult: one would have to identify books with similar sales trajectories. Moreover, granular (e.g., weekly) sales data for "similar" books are not readily available. Therefore, I employed a different strategy and dataset.

36. Rather than examine absolute changes in sales, I examined how sales rankings changed for the Works in Suit. Rankings are by their very nature relative to other books, so there is no need to collect detailed sales data for a comparison group of "similar" books. I obtained

[32] Revenue data from the AAP's Monthly StatShot reports (AAP000865 and AAP000925) support this idea for ebooks: trade ebook revenue rose from $83 million in February 2020 to $113 million in May of that year. Over the same period in the previous year, ebook revenue fell rather than rose, from $87 million in February to $82 million in May, suggesting that the rise in 2020 did not just reflect a seasonal trend.

daily sales ranking data from Amazon for hardback and paperback editions of all Works in Suit through the web service Keepa.com, which I introduced in Paragraph 11 above.[33] The data collection procedure dictates that I search for specific ISBNs, rather than a bulk download of rankings for all books. Therefore, I am limiting the analysis to a well-defined set of titles that is known to have been available through the Internet Archive's digital lending library, and for which the date of changes in availability (addition to the Internet Archive as well as removal from the Internet Archive) are well-defined: the Works in Suit. I followed the sales rankings of the print editions of these titles before and after their availability on the Internet Archive changed (see Paragraph 37, immediately below).[34] For example, I would obtain ranking and price histories for one edition of Cecilia Ahern's *PS, I Love You* by downloading the underlying information from https://keepa.com/#!product/1-0007184158.

37. For the purposes of my analysis, there are three dates on which a book's status on the Internet Archive changes: (1) the date the book was first made available for lending through the digital lending library (CDL date), (2) the date the NEL was introduced (NEL date), and (3) the date the book was removed from the digital lending library (removal date). All three dates pose a distinct shift in the nature of availability of a title at the Internet Archive, and such

[33] Amazon allows Keepa to track (daily) rankings, prices, availability through its marketplace, and ratings for a large number of products. However, Keepa does not track this information for ebooks. For that reason, I am limiting this analysis to the effects of the Internet Archive on physical book rankings. These rankings reflect an edition's rank among all physical books at Amazon. That is, they are not genre-specific.

[34] While we know for certain that the Works in Suit were available through the Internet Archive's digital lending library, we don't know whether books with similar sales rankings were available. If none of the other books were available, then the analysis provides a clean measure of the effect of the Internet Archive on Amazon sales ranking. If many of the other titles were also available, then my estimated effects will be lower bounds of the true effects. That is, the estimated effects will be closer to zero than the true effects.

distinct shifts (rather than gradual changes) help me identify the effects of such availability on the title's unit sales. While the CDL date and the removal date are title-specific (most of the Works in Suit have a removal date of June 2, 2020)[35], the NEL launched for *all titles*, including the Works in Suit, on March 24, 2020. I examined how the rankings for all print formats of a book changed around each of these dates.

38. In summary, I found no evidence that leads me to conclude that a book's addition to the Internet Archive's CDL or the launch of the NEL effected a change (in any direction) in rankings of the Works in Suit. However, I found some weak evidence of a worsening (as in, moving farther from the highest ranking of 1) of a book's Amazon ranking after the title was no longer available for borrowing from the Internet Archive. Below, I explain how I arrived at these findings.

39. Data collection: The data collection consists of two steps: (1) For each of the Works in Suit, I created a list of ISBNs for all its print editions;[36] (2) For each of these editions, I collected daily Amazon ranking, price, and rating histories from Keepa. To perform the first step, I used Goodreads to obtain all editions of each title.[37] For example, for "Gone Girl" I obtained all ISBNs on the Goodreads website, available at:

---

[35] INTARC00474639.

[36] Recall that I am only looking at print editions because ranking histories for electronic versions are not available. Because print sales continue to outweigh ebook sales by a large margin (see Paragraphs 16–19 above), the analysis still captures a large part of each title's sales.

[37] Goodreads is a social book cataloging website in which users can register and review books. The site has millions of members. By 2012, it had already shelved 395 million books. See "These are Top 25 Book Reviewers on Goodreads," *Forbes*, available at https://www.forbes.com/sites/markfidelman/2012/10/16/goodreads-ceo-these-top-25-book-reviewers-represent-the-future-infographic/, (Oct. 16, 2012). Goodreads thus provides a good starting point for creating a list of ISBNs for a set of titles.

https://www.goodreads.com/work/editions/13306276?utf8=%E2%9C%93&per_page=100. I recorded the ISBNs of all hardcover and paperback editions, dropping editions of other formats because daily ranking data are not available for them. Then, to perform the second step, I used Keepa's API to search for and collect ranking histories for the remaining ISBNs. The API allows me to search for each individual ISBN at a time, and it provides ranking histories going back in time as far as to March 2011 for some editions. I found ranking and other information for a total of 1,260 editions for 118 titles.[38]

40. Potential shortcomings of the data: The resulting dataset provides daily sales rank information on Amazon, which allows me to examine the immediate and medium-term effects of changes in availability at the Internet Archive well, but the nature of this dataset has some potential shortcomings. First, the dataset only captures part of the market. Whereas the dataset is almost comprehensive for hardcover and paperback versions, it does not capture electronic books. Thus, any interpretation of the results should be limited to print books.

41. Second, the dataset only captures sales through Amazon, not through other channels. Media and technology analyst Benedict Evans reports that Amazon had at least a 50% market share for print books in 2019.[39] If sales for the Works in Suit—relative to other books—did not evolve differently at Amazon than at other outlets, this limitation does not affect the results.

42. Third, the dataset reports daily sales rankings rather than unit sales. In my analysis (described below), I work with these ranking data because they are "true" in the sense

[38] Keepa does not track all editions and titles, so that the dataset does not include the full set of 127 Works in Suit.

[39] See "What's Amazon's market share?," *Benedict Evans*, available at https://www.ben-evans.com/benedictevans/2019/12/amazons-market-share19, (Dec. 19, 2019).

that they are not approximated. However, when I quantify the estimated effects of availability at the Internet Archive, I will translate these rankings into "pseudo-sales" data. To do this, I use a formula proposed by Reimers and Waldfogel (2021), who created such a translation based on information about weekly Top 100 print bestsellers in 2018 from NPD Bookscan along with Amazon ranking data for some of those editions. Their formula is: unit sales = 10,167 * rank^(-0.45).[40] This formula suggests a relatively steep drop in sales as a book's ranking deteriorates (moved farther away from the highest ranking of 1). If, for example, a book's ranking on a given day at Amazon is 10, then the formula suggests 10,167 * 10^(-0.45) = 3607 units of the book are sold on that day. For a book that is ranked 100th at Amazon, I approximate its unit sales on that day to be 1,280.

43. Analysis of the effects of availability on Amazon sales rankings: Amazon's rankings do not represent a snapshot of sales on one day but rather represent the sum of book sales from a previous period of several (unspecified) days. That is, rankings on Amazon are based on a *moving average* of book sales in the past. In my analysis, I account for this moving-average property. Specifically, I control for the effects of past rankings, which allows me to focus on the effect of Internet Archive availability on sales on a particular day. The econometric method I use for this is a regression analysis, in which I regress the natural log of a book's Amazon ranking on a specific day on information about its concurrent availability on the Internet

[40] See page 1957 of Reimers, Imke & Joel Waldfogel. "Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings." *American Economic Review* 111.6 (2021): 1944–71. The main assumption for arriving at the formula is that the relationship between Amazon rankings and sales can be described by a power law, whereby a relative change in one ranking is associated with a proportional relative change in sales, independent of the initial rank. In other words, sales vary as a power of rankings. This is a common assumption in the academic literature when describing the relationship between rankings and sales.

Archive (this is the variable that I am most interested in) as well as its previous ranking (to account for the fact that a book's ranking depends on its previous ranking as well as its concurrent sales) and a polynomial function of the book's age (the number of months since its publication).[41]

44. Broadly speaking, my analysis allows me to capture the effect of changes in a book's availability through the Internet Archive on the Amazon sales ranking of its editions. The regression goes beyond a simple correlation between availability and rankings. Rather, it is a *conditional* correlation: I include information about the edition's age as well as the identity of the title in order to condition on these variables. For example, for the October 2009 hardcover edition of *What the Dog Saw*, I compare its Amazon sales ranking shortly before and shortly after the book becomes available at the Internet Archive. That is, instead of comparing sales rankings across different books, I can focus in on each individual book as its availability at the Internet Archive's lending library changes.

45. Results – Controlled Digital Lending: To estimate the effect of changing availability of a Work in Suit on the sales ranking of its print editions, I obtained—for all Works in Suit—the date on which any version of the title first was made available through CDL through

[41] Formally, I estimate the following regressions equation with daily data for all editions of the Works in Suit: $\ln(rank_{jt}) = \theta \ln(rank_{jt-1}) + f(time\ since\ pub_{jt}) + \beta\ afterIA_{jt} + v_i + \epsilon_{jt}$, where $afterIA$ indicates whether the title has been made available at the Internet Archive, and $f(time\ since\ pub_{jt})$ describes a 3rd order polynomial: $\alpha_1 time + \alpha_2 time^2 + \alpha_3 time^3$. The regression equation I am using here closely follows Reimers, I. & Waldfogel, J. "Digitization and Pre-Purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings," *American Economic Review,* 111.6 (2021): 1944–71. The coefficient β describes the conditional correlation between a title's availability at the Internet Archive and the natural log of its ranking. Mathematically, this allows me to create a connection between an edition's change in availability on the Internet Archive and a percentage change in its ranking, through: % change in ranking = $(e^{\beta} - 1) \times 100\%$.

the Internet Archive's lending library.[42] I then estimated a regression of the form described in the previous two paragraphs, using daily ranking data of all editions for the Works in Suit. Based on these regressions, I estimated the coefficient β from footnote 41 (the coefficient that describes the conditional relationship between Internet Archive availability and the Amazon sales rank) to be $\beta = +0.0024$. Taking the estimated coefficient at face value, the regression implies that the book's rank increases (worsens) by about 0.24%. However, as I explain below, the result is not statistically significant.[43]

46. Because this result is based on a sample of books, rather than the universe of all books, this coefficient is estimated with some uncertainty. I can express this uncertainty with the standard error of the coefficient, which is 0.0017. The coefficient and standard error together provide information about the statistical significance of the coefficient. Roughly speaking, if the standard error (the square root of the variance) is large compared to the size of the coefficient, then we cannot say with certainty that the coefficient is different from zero. In other words, we cannot say with certainty that there is a relationship between a change in a book's availability at the Internet Archive and the Amazon sales rankings of its print editions because a coefficient of zero indicates no effect. To illustrate the statistical significance of the coefficients, I will use 95% confidence intervals. My coefficient and standard error imply that, if one were to estimate

[42] INTARC00474639.

[43] In addition, the estimated coefficient varies depending on the set of data that I include and the choices of control variables. Here I present a "conservative" estimate that is based on all available ranking data. In Exhibit C, I present a Table that shows the estimated coefficients using a wide range of included time periods around each edition's CDL date. In Exhibit D, I also present the estimated coefficients from regressions that use (1) three months before and after the change in availability, and (2) one month before and two months after the change. Exhibit D also includes the estimated coefficients of regressions studying the other two changes in availability (NEL, and removal from the Internet Archive). In Paragraph 46 below, I provide a more flexible specification that removes some of the reliance on assumptions.

this regression 100 times (on 100 different samples), 95 of the 100 coefficients are expected to range between -0.0010 (for a rank improvement of 0.10%) and +0.0057 (for a worsening of 0.57%).[44] Because this interval includes zero, I cannot statistically reject the possibility that availability at the Internet Archive has no effect on a book's ranking, and thus the result is not statistically significant.

47. Any effects of a change in a book's availability through the Internet Archive on Amazon print book rankings may vary over time. For example, knowledge about the book's availability on the Internet Archive could spread slowly so that potential readers only learn about it several weeks after the book's digitization and availability for borrowing. I test this in Figure 3. Rather than aggregating all weeks before and after a title became available at the Internet Archive, I now examine changes in a book's ranking in each week before and after its inclusion in the Internet Archive. I present the estimated differences in a book's Amazon ranking in each week (from 20 weeks before inclusion to 30 weeks after it) relative to the week before the title became available, in Figure 3. In short, the figure shows no statistically significant effect of CDL on rankings. Rather, there may be larger changes in rankings associated with events other than the book's inclusion at the Internet Archive.[45] Because the

[44] I choose 95% confidence intervals, which describes whether a coefficient is statistically significant at the 5% level, because this is the standard cutoff for statistical significance in the economics literature. If the confidence interval is relatively narrow (in other words, the range of calculated numbers is small), the coefficient is precisely estimated. If that interval includes zero, we cannot reject the hypothesis that the true coefficient is zero.

[45] Note that the reversed y-axis captures the inverse relationship between rankings (where low values—close to 1—are good) and sales (where high values are good). As a rule of thumb, in all pictures, values toward the top of the graph imply better rankings (higher sales), and values toward the bottom of the graph represent worse rankings (lower sales).

initial change in Amazon rankings occurred before the change in availability, I cannot connect these ranking changes cleanly to the title's addition to CDL.

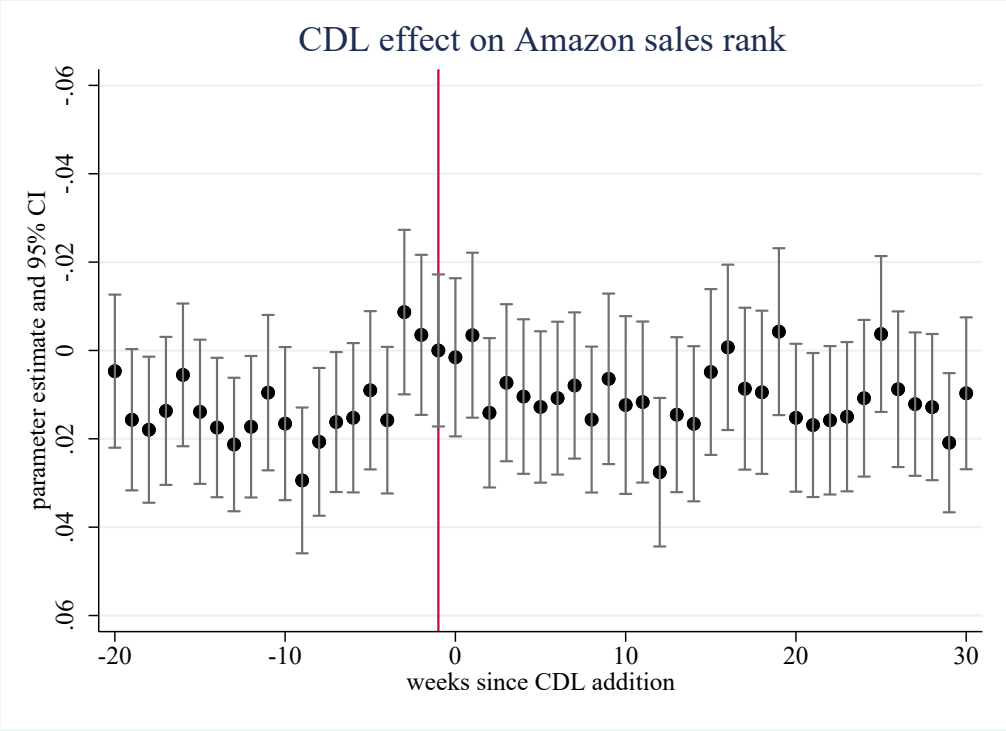

Figure 3: Estimated Change in Sales Rank due to the Internet Archive's CDL, for each week before and after the book's addition

48. Results – Removal from the Internet Archive: The filing of this lawsuit provided another distinct change in the availability of the Works in Suit on the Internet Archive: many of the Works in Suit were removed from all lending at the Internet Archive on June 2, 2020 (before the NEL terminated), and some were removed shortly after the termination of the NEL. The removal of the books from the Internet Archive was arguably not driven by immediate changes in the demand for these titles but rather a response to the filed suit. Therefore I can study the effects of this removal of a title on the Amazon sales rankings of its print editions directly.

49. To study these changes, I repeat the analysis from just above. In Figure 4, I plot the estimated differences in a book's Amazon ranking in each week around the book's removal

from the Internet Archive, relative to the week immediately preceding its removal.[46] The figure suggests two things. First, the coefficients seem to increase slightly leading up to the book's removal from the Internet Archive (red line). This suggests a possibility that the Works in Suit had slightly more sales, relative to other titles, leading up to their removal. While this period coincides with the titles' availability, I do not have enough information to connect the relative sales increase to the NEL. Second, the estimated rankings drop (worsen) almost immediately after the book's removal from the Internet Archive. That is, when the books are no longer available on the Internet Archive, their unit sales at Amazon decrease relative to other books. Regression analyses based on the regression equation in Footnote 41 suggest a worsening of the Amazon rank between 1% and 2% when the work is taken down, with relatively narrow confidence intervals, especially as I consider narrow time frames (a few weeks) around a work's removal. I show the estimated coefficients from two such regressions in Exhibit D. Taken at face value, this is at least suggestive evidence that availability of a book at the Internet Archive does not hurt its sales.

[46] This figure is analogous to Figure 3 above, using the weeks around a book's removal from the Internet Archive instead of the weeks around its addition to CDL.

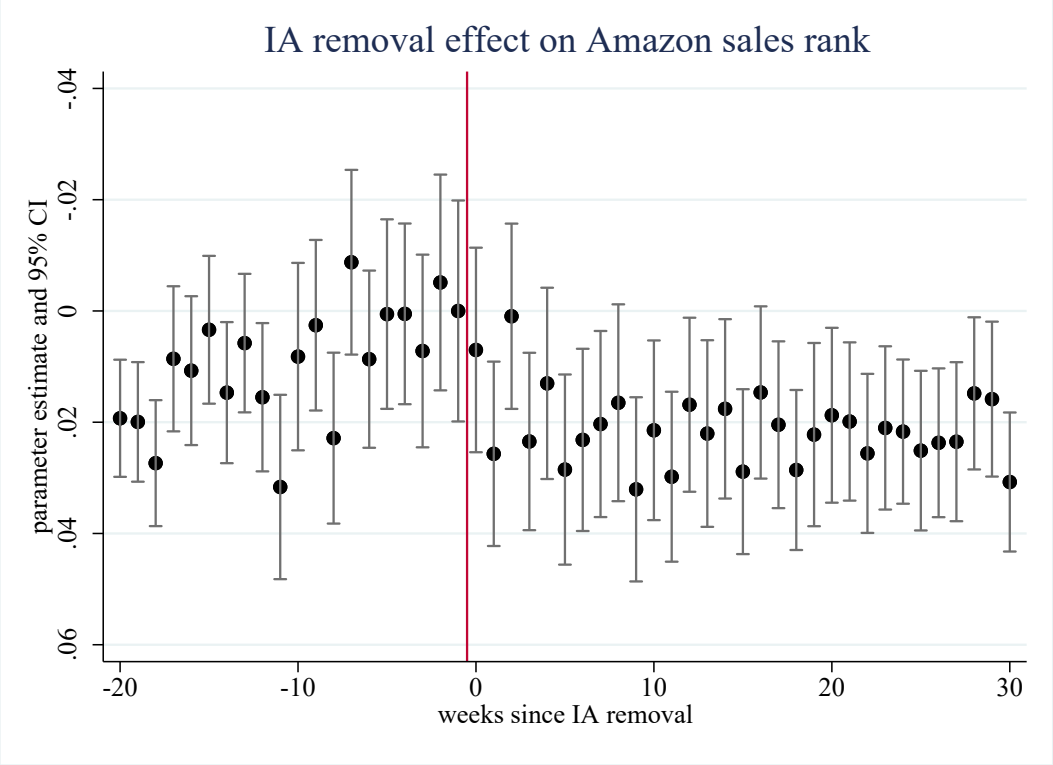

Figure 4: Estimated Change in Sales Rank due to removal from the Internet Archive's CDL, for each week before and after the book's removal

50. There is a possibility that this estimated negative effect of the removal from the Internet Archive is driven by seasonal patterns. Most of the Works in Suit were taken off the Internet Archive in June 2020. If the Works in Suit naturally experienced a drop in their sales around the beginning of summer, then the coefficients I estimated would simply reflect a seasonal change in their sales (relative to other books), rather than an effect of the Internet Archive. To test whether the effect of the removal that I estimated above was in fact driven by seasonal changes in the works' readership, I perform a "placebo" analysis: I pretend that each book's removal from the Internet Archive ("treatment") happened exactly one year before its true removal date, and I examine changes in Amazon sales ranks for the Works in Suit around

that "placebo treatment" date.[47] Figure 5 shows the estimated changes in Amazon sales rankings. Comparing Figures 4 and 5 shows that sales rankings worsened more after the true removal (Figure 4) than after the placebo removal a year prior to that date (Figure 5), although a worsening of the ranking cannot be ruled out after the placebo removal either. This suggests that the true removal indeed had an effect on Amazon sales rankings.

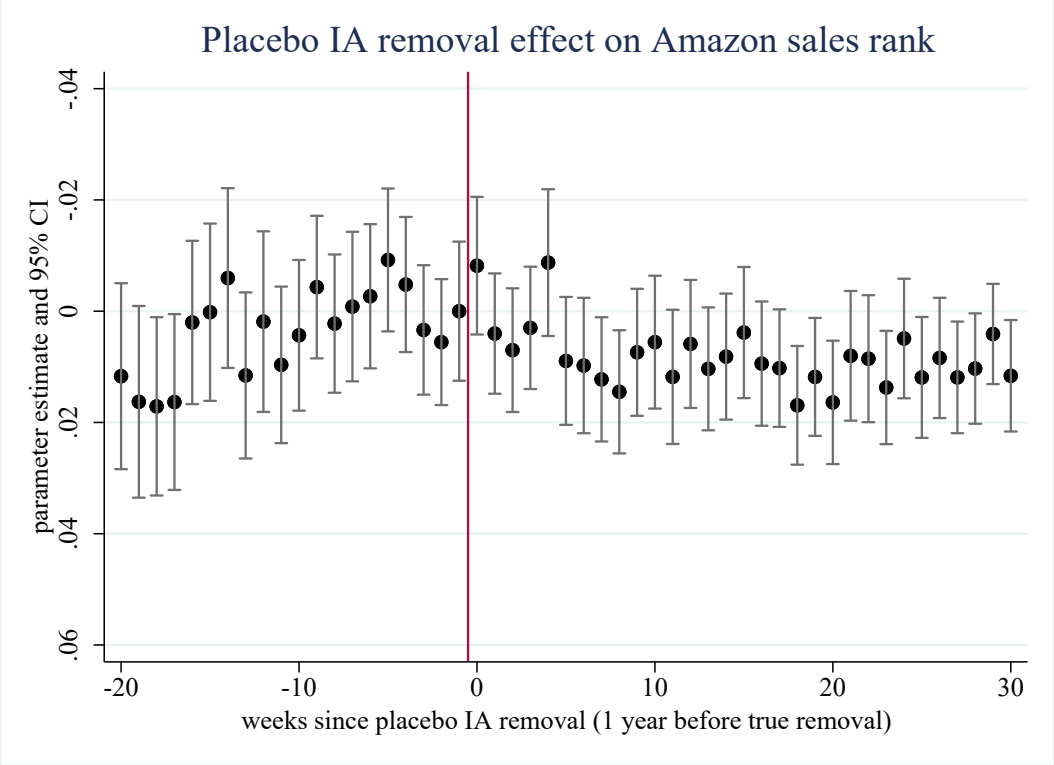

Figure 5: Estimated Change in Sales Rank due to placebo removal from the Internet Archive's CDL, where the placebo removal happened exactly one year prior to the book's true removal

51. Results – NEL: The third change in a book's online availability happened when the NEL was launched. The NEL launched on March 24, 2020, making books even more accessible on the Internet Archive than before by temporarily lifting the owned-to-loaned waitlist

[47] Note that I do not perform a placebo analysis for the effects of a book's inclusion in CDL because placebo tests are generally performed to examine whether an estimated effect is due to chance, but I found no effect of CDL to begin with.

parameter of CDL. Although this change in a book's availability at the Internet Archive can be seen as relatively small, it is possible that the improved availability during the NEL had a distinct effect on the unit sales of print editions. I examined this possibility using the same framework as above, but instead of examining how the sales rank changed after the entire addition to or removal from the Internet Archive, I now examine how the sales rank for the Works in Suit changed when they became more readily available, due to the NEL.

52. Again, I examined changes in the Amazon ranking of the Works in Suit in each week around the launch of the NEL. I present the estimated ranking differences in each week (from 20 weeks before launch to 30 weeks after it), relative to the week before the NEL launch, in Figure 6. The figure is analogous to Figures 3 and 4, with a different focal date. It suggests that, whereas sales rankings of the Works in Suit did not trend upward or downward before the NEL launch (the first vertical line in the chart), their rankings improved significantly—by about 2%—almost immediately after the launch, and rankings remained improved throughout their NEL tenure. After about 12 weeks—around the time that most Works in Suit were removed from the Internet Archive's lending library (June 2, 2020)—their Amazon rankings seemed to revert toward their original levels within a few weeks.

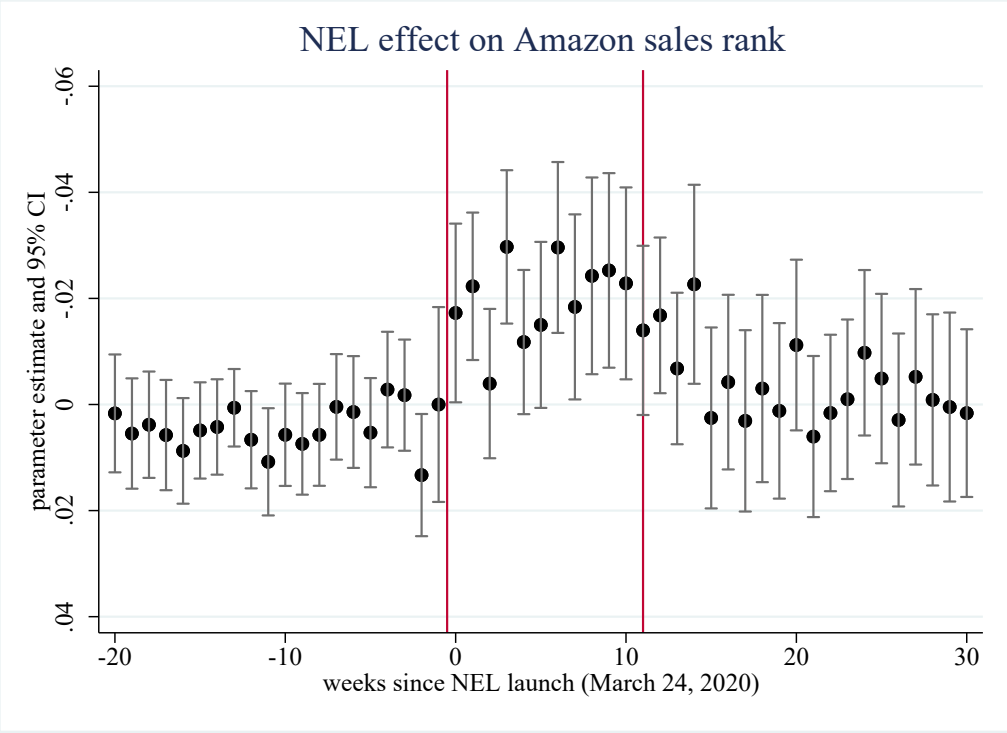

Figure 6: Estimated Change in Sales Rank due to the Internet Archive's NEL, for each week around its launch. The first red line indicates the launch of the NEL, the second line indicates its shutdown.

53. The improvement in the rankings for the works in suit could be due to the NEL, but they might also be due to seasonal effects. Here, I repeat the placebo test from Figure 5. In particular, I impose a "placebo" NEL launch date one year before the true launch, and I examine how Amazon sales rankings for the Works in Suit change around this placebo date (March 24, 2019). If there are no seasonal trends in the rankings of the Works in Suit around springtime, then the estimated coefficients the placebo launch should be close to zero. However, Figure 7 shows that the estimated coefficients—and therefore the estimated rankings for the Works in Suit—also improved in March of 2019, with size effects similar to the improvements observed in March 2020. Thus, because Figures 6 and 7 look quite similar to each other, I cannot conclude that the NEL had either a positive or a negative effect on the Amazon sales rankings of print editions.

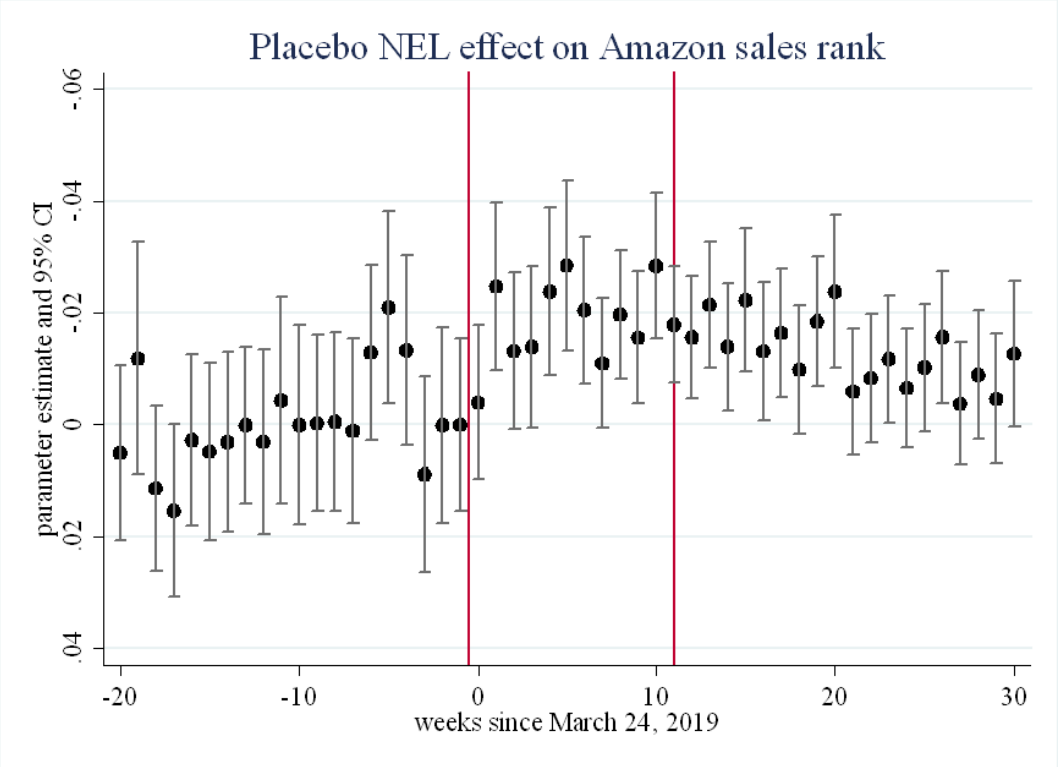

Figure 7: Estimated Change in Sales Rank due to the Internet Archive's NEL, for each week around its launch. The first red line indicates the launch of the NEL, the second line indicates its shutdown.

54. The size of the effect relative to lifetime sales: In the above paragraphs, I document no clear change in Amazon rankings for the Works in Suit as they become available through CDL (CDL date) or as their availability improves further with the launch of the NEL. By contrast, I present weak evidence of a small but statistically significant worsening of the Amazon ranking when a book is removed from the Internet Archive. Note, however, that the vast majority of books are made available on the Internet Archive five years or more after their original publication. Based on my analysis in Section B, it is unlikely that either positive or negative effects related to a title's availability on the Internet Archive account for a large share of a book's lifetime sales. For example, if only 10% of a book's lifetime sales typically occur later than the first five years after its publication (as is suggested by my analysis of lifetime sales above), then a (hypothetical, permanent) 2% change in sales after 5 years would indicate only a

0.2% change in the book's lifetime sales. But, as noted above, there is no evidence to suggest that the availability of the Works in Suit on the Internet Archive adversely affected their sales.

55. Conclusions:

a. The share of ebook sales has decreased steadily since 2014, but there is no evidence that this decline over the last five to ten years is related to the Internet Archive's digital lending program.

b. Sales of an edition of a title tend to decrease rapidly after the first year. Sales in the first five years after an edition's publication account for up to 90% of its lifetime sales. Most of the Works in Suit behaved consistently with this observed pattern.

c. My analysis of the effects of availability at the Internet Archive on Amazon sales rankings of print versions of the Works in Suit leads me to draw two conclusions. First, I find no statistically significant evidence that a book's inclusion on the Internet Archive's lending library or the launch of the NEL hurt or harmed its sales on Amazon. Second, I find some weak evidence that a book's removal from the Internet Archive is associated with lower (worse) rankings of print editions at Amazon.

Dated: February 25, 2022

DocuSigned by:
Imke Reimers, Ph.D.
6A0C28ABD670450...

IMKE REIMERS, PH.D.

Exhibit A

Curriculum Vitae
Spring 2022

IMKE C. REIMERS

Academic Appointments:

Associate Professor (with tenure), Department of Economics, Northeastern University, Summer 2020 – present
Fellow, Center for Advanced Studies, Ludwig Maximilian University Munich – Spring 2021
Assistant Professor, Department of Economics, Northeastern University, Fall 2014 – Spring 2020
Postdoctoral Research Fellow, National Bureau of Economic Research, Fall 2013 – Summer 2014

Faculty Affiliate, Digital, Analytics, Technology and Automation (DATA) Initiative, Northeastern University

Education:

Degree	*Field*	*Institution*	*Year*
Ph.D.	Economics	University of Minnesota	2013
M.A.	Economics	University of Minnesota	2012
B.Sc.	Mathematics and Economics	University of Nebraska (high distinction)	2008

Publications

"Digitization, Prediction and Market Efficiency: Evidence from Book Publishing Deals" with Christian Peukert, *Management Science* (2022)

"Digitization and Pre-Purchase Information: The Causal and Welfare Effects of Reviews and Crowd-Based Ratings" with Joel Waldfogel, *American Economic Review,* 111.6 (2021), 1944-71

"The Impacts of Telematics on Competition and Consumer Behavior in Insurance" with Benjamin Shiller, *Journal of Law and Economics*, 62.4 (2019), 613-32

"Copyright and Generic Entry in Book Publishing," *American Economic Journal: Microeconomics,* 11.3 (2019), 257-84

"Do Coupons Expand or Cannibalize Revenue? Evidence from an E-market" with Claire (Chunying) Xie, *Management Science*, 65.1 (2018): 286-300.

"Examining Regulatory Capture: Evidence from the NHL" with Gregory DeAngelo and Adam Nowak, *Contemporary Economic Policy*, 36.1 (2018), 183-91

"Are Public and Private Enforcement Complements or Substitutes? Evidence from High Frequency Data" with Gregory DeAngelo and Brad Humphreys, *Journal of Economic Behavior and Organization*, 141 (2017): 151-63

"Throwing the Books at Them: Amazon's Puzzling Long-Run Pricing Strategy" with Joel Waldfogel, *Southern Economic Journal*, 83.4 (2017): 869-85 (runner-up, Georgescu-Roegen Prize for Best Article in the Southern Economic Journal 2017.)

"Can Private Copyright Protection be Effective? Evidence from Book Publishing" *Journal of Law and Economics*, 59.2 (2016): 411-440

"Storming the Gatekeepers: Digital Disintermediation in the Market for Books" with Joel Waldfogel, *Information Economics and Policy*, 31 (2015): 47-58

Select Working Papers

"Digitization and the Demand for Physical Works: Evidence from the Google Books Project" with Abhishek Nagaraj

"Does Amazon Exercise its Market Power? Evidence from Toys R Us" with Leshui He and Benjamin Shiller

"Visibility of Technology and Cumulative Innovation: Evidence from Trade Secrets Laws" with Bernhard Ganglmair

"The Digital Challenge to Public Libraries" with Joel Waldfogel

"Home Sweet Home? Covid-19, Stadium Attendance and Efficiency in Sports Betting Markets" with James Dana

Invited Seminars and Conference Presentations (including scheduled)

2022 Harvard Law School; Boston University; University of Toronto; Stanford University; Paris Seminar on the Economics of Digitization; University of Colorado

2021 Rotterdam School of Management; Harvard Business School; University of Pittsburgh; ZEW Mannheim; Digital Economy Workshop, Munich, Germany (panelist and discussant); International Industrial Organization Conference; Bureau of Labor Statistics; ZEW Conference on Information and Communication Technologies (discussant); Department of Justice; Georgetown University; National University Singapore; Brandeis University; OECD; RISE4 Workshop, Max Planck Institute Munich (discussant); University of East Anglia; University of Giessen

2020 13th Digital Economics Conference, Toulouse School of Economics, Toulouse, France; Boston College (canceled); 8th Annual Intellectual Property Law Conference, Austin, TX (canceled); University of Colorado, Boulder (canceled); NBER Summer Institute Digitization Meeting; IP Day, Technology & Policy Research Institute, Boston University; 13th Annual USPTO-Northwestern Innovation Economics Conference (discussant); Technology & Declining Economic Dynamism, Technology & Policy Research Institute, Boston University; Ringvorlesung, Department of Economics, University of Mannheim; CAS Research Group, LMU Munich, Germany; MaCCI-ZEW Workshop on Economics of Innovation, Mannheim, Germany (discussant)

2019 AEA Meetings, Atlanta, GA; 12th Digital Economics Conference, Toulouse School of Economics, Toulouse, France; NBER Economics of Digitization Winter Meeting, Palo Alto, CA; Lafayette College; International Industrial Organization Conference, Boston, MA; 8th ZEW Conference on the Economics of Innovation and Patenting, Mannheim, Germany; Munich Summer Institute, Munich, Germany; NBER Summer Institute, Innovation Meeting; Tufts University; Pennsylvania State University; Queen's University; Workshop on the Economics of Fixed Book Price Systems, Giessen, Germany; University of Luxembourg

2018 11th Digital Economics Conference, Toulouse School of Economics, Toulouse, France; NBER productivity lunch, Cambridge, MA; NBER Economics of Digitization Winter Meeting, Palo Alto, CA; International Industrial Organization Conference, Indianapolis, IN; Roundtable on Empirical Methods in Intellectual Property, Northwestern Pritzker School of Law, Chicago, IL; Munich Summer Institute, Munich, Germany; Society for Economic Research on Copyright Issues, Toronto, Canada; Platform Strategy Research Symposium, Boston University, Boston, MA (discussant); NBER Summer Institute (poster); Bates College; Roundtable on Copyright Issues, Notre Dame University, South Bend, IN; National Association for Business Economics TEC, San Francisco, CA; University of Connecticut

2017 Dartmouth Winter IO Conference, Hanover, NH; Copyright and Technology Conference, New York, NY; NBER productivity lunch, Cambridge, MA; MIT Press, Cambridge, MA; International Industrial Organization Conference, Boston, MA; Harvard Business School Workshop on Innovation in a Global Economy, Boston, MA; 8th SEARLE Conference on Internet Commerce and Innovation, Chicago, IL; ZEW Conference on Information and Communication Technologies, Mannheim, Germany; Western Economic Association Annual Meeting, San Diego, CA

2016 AEA Meetings, San Francisco, CA; Annual Scientific Seminar on Competition and Regulation in Infrastructure and Digital Markets, Florence, Italy; International Industrial Organization Conference, Philadelphia, PA; Rijksuniversiteit Groningen; Southern Economic Association Annual Meeting, Washington, DC.

2015 NBER productivity lunch; International Industrial Organization Conference, Boston; Strategic Business Management & Economic Research Conference, Boston; ZEW Conference on Information and Communication Technologies, Mannheim, Germany; NBER Summer Institute (poster); University of Calgary Empirical Microeconomics Workshop, Banff, AB, Canada; Southern Economic Association Annual Meeting, New Orleans, LA

2014 NBER productivity lunch; Eastern Economic Association Annual Meetings, Boston, MA; Workshop on Digital Media Markets and the Modernization of Copyright in the EU, Brussels, Belgium; International Industrial Organization Conference, Chicago, IL; Rijksuniversiteit Groningen, ZEW Conference on Information and Communication Technologies, Mannheim, Germany; North American Summer Meetings of the Econometric Society, Minneapolis, MN; Frankfurt Book Fair; Workshop on the Interaction Between Legal and Pirated Book Sales, Seville, Spain

2013 Kansas State University, Rensselaer Polytechnic Institute, Northeastern University, Dartmouth College, Rijksuniversiteit Groningen, University of Bremen, Massachusetts Institute of Technology

Honors and Awards

Finalist, College of Social Sciences and Humanities Outstanding Teaching Award, Northeastern University, 2020
Postdoctoral Fellowship, Digitization and Copyright Initiative, National Bureau of Economic Research. 2013
Graduate Research Program Partnership Dissertation Fellowship, University of Minnesota. 2012
Distinguished Instructor, Department of Economics, University of Minnesota. 2010
Distinguished Teaching Assistant, Department of Economics, University of Minnesota. 2009, 2010
Silverman Fellowship, Department of Economics, University of Minnesota. 2008-2009
NCAA Postgraduate Fellowship, National Collegiate Athletic Association 2008
Phi Beta Kappa, Gamma Chapter of University of Nebraska, Lincoln, Nebraska 2008 – present

Teaching

Northeastern University
Microeconomic Theory (undergraduate)
Applied Econometrics (undergraduate)
Framework of Industrial Organization (Ph.D.)

University of Minnesota:
Industrial Organization and Antitrust Policy (undergraduate)
Principles of Macroeconomics (undergraduate)
Applied Econometrics (Ph.D., teaching assistant)
Principles of Microeconomics (undergraduate, teaching assistant)
Development Economics (undergraduate, teaching assistant)

Guest lectures:
Massachusetts Institute of Technology, HEC Lausanne, Brandeis University, Northeastern (Network Sciences), Suffolk University

Advising

PhD (all at Northeastern University, committee member unless otherwise specified):

Diana Li (expected 2024, main advisor)
Yanli Liu (expected 2023, main advisor)
Philip Rubin-Streit (expected 2023, main advisor)
Yanchi Zou (expected 2023)
Andrew Kearns (expected 2022, co-main advisor)
Farzaneh Nekui (expected 2022)
Arvind Sharma (expected 2022)
Yunus Yilmas (2021, Economist, Analysis Group)
Tuan Nguyen (2021, Economist, RSM US)
Richeng Piao (2021, Lecturer, Northeastern University)
Robert Bradley (2020, Head of Research, State Street)
Carlos Casso Dominguez, (2020, Lecturer, Boston University)
Shuang Wang (2020, Economist, Amazon)
Joanna Fister (2019, Federal Communications Commission)
Ngoc Ngo (2019, Competition Dynamics)
He Wang (2019, Assistant Professor, Beijing Sport University)
Hanchun Zhang (2019, Researcher at Harvard Business School)
Pukar KC (2018, JP Morgan Chase)
Nehan Naim (2018, Assistant Professor, College of Environmental Science and Forestry, SUNY Syracuse)
Lihua Han (2016, Discover Financial Services)

MA thesis advisor (all at Northeastern University):
Xiaonan Qin, Sinan Inan

Undergraduate thesis supervisor (all at Northeastern University):
Yi Fang, Ryan Megahey, Paul Molander, Andrew Wehner

Other Service activities

Ad hoc referee:
American Economic Journal: Policy; American Economic Review; Information Systems Research; International Journal of Industrial Organization; Journal of Economics and Management Strategy; Journal of Industrial Economics; Journal of Law, Economics and Organization; Journal of Law and Economics; Journal of Political Economy; Journal of Public Economics; Management Science; RAND Journal of Economics; Review of Economics and Statistics; Quantitative Marketing and Economics; Economic Inquiry; Economics; Information Economics and Policy; International Journal of Economics and Business; International Journal of Services and Operations Management; Journal of Cultural Economics; Journal of Regulatory Economics; Journal of Media Economics; Journal of Risk and Insurance; Oxford Economic Papers; Review of International Economics; Southern Economic Journal

Other:
At Northeastern:

Member, faculty search committee, Northeastern University, 2013 – 2014
Co-organizer, Economics Department Seminar, Northeastern University, 2014 – 2020
Member, CSSH Collaborative Research Award Committee, Northeastern University, 2020
Member, PhD admissions committee, Northeastern University, 2021 – 2022

Profession:

Guest Co-Editor, Information Economics and Policy, Special Issue on catastrophic events, 2021
Member, Program Committee, 20th International Industrial Organization Conference, 2022
Member, Scientific Committee, 20th ZEW Conference on the Economics of Information and Communication Technologies, 2022
Editorial Board, Marketing Science, 2022-present

Exhibit C

Table 1: Estimated coefficients from regressions of a book's Amazon sales rank on its availability through CDL (varying time windows)

	(1) 10 days	(2) 20 days	(3) 30 days	(4) 60 days	(5) 90 days	(6) 180 days	(7) 365 days
Post CDL	-0.00285	-0.0111	0.00846	0.0183**	0.00923	0.00196	0.000843
	(0.0195)	(0.0142)	(0.0117)	(0.00835)	(0.00685)	(0.00487)	(0.00349)
Lagged log-rank	0.475***	0.572***	0.609***	0.684***	0.709***	0.736***	0.760***
	(0.0128)	(0.00843)	(0.00654)	(0.00421)	(0.00333)	(0.00225)	(0.00153)
Edition age (mos)	0.0402	0.0750***	0.0174	-0.000743	0.00325	0.00489***	0.00537***
	(0.0684)	(0.0257)	(0.0141)	(0.00519)	(0.00284)	(0.00102)	(0.000378)
(Edition age)2	-7.88e-05	-0.000158	-4.02e-05	-4.46e-06	-1.15e-05	-6.25e-06	-9.87e-06***
	(0.000422)	(0.000160)	(8.65e-05)	(3.26e-05)	(1.78e-05)	(6.33e-06)	(2.32e-06)
(Edition age)3	-5.53e-07	-1.91e-07	-6.62e-08	-1.73e-08	2.22e-08	4.17e-09	6.32e-09***
	(4.08e-07)	(1.60e-07)	(8.98e-08)	(3.95e-08)	(1.99e-08)	(6.32e-09)	(2.30e-09)
Observations	5,216	10,035	15,121	30,195	44,793	90,395	178,753
R-squared	0.986	0.985	0.984	0.984	0.983	0.983	0.982

Notes: Coefficients from regressions of log-rankings on the post-CDL indicator and control variables. All regressions include fixed effects (indicator variables) for each edition. Column headings indicate the number of days around the change in the edition's CDL availability that are included in the regression. Standard errors in parentheses. Statistical significance is indicated by stars: *** p<0.01, ** p<0.05, * p<0.1

Exhibit D

Table 2: Estimated coefficients from regressions of a book's Amazon sales rank on its availability at the Internet Archive

	CDL		IA removal		NEL	
	(1)	(2)	(3)	(4)	(5)	(6)
	3 mos	1/2 mos	3 mos	1/2 mos	3 mos	1/2 mos
Post CDL	0.00923	0.0116				
	(0.00685)	(0.00893)				
Post IA removal			0.0274***	0.0120		
			(0.00657)	(0.00859)		
Post NEL					-0.0266***	-0.0230***
					(0.00552)	(0.00811)
Lagged log-rank	0.709***	0.649***	0.706***	0.617***	0.734***	0.697***
	(0.00333)	(0.00515)	(0.00310)	(0.00497)	(0.00254)	(0.00419)
Edition age (mos)	0.00325	0.00703	0.0178***	0.0232***	-0.00146	-0.00769
	(0.00284)	(0.00490)	(0.00267)	(0.00459)	(0.00217)	(0.00652)
(Edition age)2	-1.15e-05		-7.46e-05***		1.13e-05	3.97e-05
	(1.78e-05)		(1.08e-05)		(8.96e-06)	(2.77e-05)
(Edition age)3	2.22e-08		6.09e-08***		-2.51e-08**	-6.91e-08**
	(1.99e-08)		(1.26e-08)		(1.03e-08)	(3.18e-08)
Observations	44,793	22,244	52,209	25,374	71,923	29,839
R-squared	0.983	0.984	0.980	0.978	0.984	0.984

Notes: Coefficients from regressions of log-rankings on the post-CDL indicator and control variables. All regressions include fixed effects (indicator variables) for each edition. The first two columns report estimated effects of the title's inclusion in CDL; the next two columns show estimated effects of the title's removal from the Internet Archive; the last two columns show estimated effects of the NEL. Odd-numbered columns use 3-month windows around the focal date, and even-numbered columns use the one month before and two months after the focal date. Standard errors in parentheses. Statistical significance is indicated by stars: *** p<0.01, ** p<0.05, * p<0